

NETAPP AI Pod



Introducing NetApp® AI Pod™, an AI-focused converged infrastructure built on NetApp ONTAP® all-flash storage

Artificial intelligence (AI) is a transformative technology that is reshaping business across industries, offering vast opportunities for innovation and efficiency. The adoption of AI and generative AI (GenAI) enables businesses to automate tasks, predict trends, personalize customer experiences, and optimize processes, resulting in increased productivity and reduced costs. However, to realize the potential of AI, organizations need a strategic approach to adoption, including efficient AI deployments, robust data governance and security, and effective collaboration among IT operations, data scientists, and developers.

AI is the biggest disruptive force of this era, and GenAI is a tidal wave. According to [Forrester](#), enterprise AI initiatives can boost productivity and creative problem-solving by 50%. Moreover, up to 60% of workers are expected to use their own AI to perform their tasks in the coming years. The transformative power of AI and GenAI has the potential to revolutionize enterprise operations

Business challenges in adopting AI

Complex integrations and siloed infrastructures are among the top operational obstacles when adopting AI. Numerous components need to be integrated in an AI project, including compute resources (like CPUs and GPUs), storage systems, networking devices, and various software components (like machine learning [ML] libraries, data processing tools, and analytics platforms). Each of these components can come from a different vendor, can have unique requirements, and might need to be configured in specific ways to work together.

Integrating these disparate components requires deep technical expertise and a good understanding of how each component works. If the integration is not done correctly, it can result in performance issues, system instability, and other problems that can slow down the AI project and waste valuable resources.

Without the baseline of a unified strategy for infrastructure, different teams or departments may set up their own infrastructure for their specific AI projects. However, these silos can cause inefficient use of resources. For example, one team's infrastructure might be underused while another team's infrastructure is overloaded. Silos can also hinder collaboration and knowledge sharing, complicate data management, and make it harder to scale AI projects. As the demand for AI resources grows, each silo might need to be scaled separately, which can be more complex and costly than scaling a unified infrastructure.

AI workloads often involve complex computations that require significant processing power. If the underlying infrastructure is not properly configured, unpredictable performance can result. For instance, there might be delays in processing, slow response times, or even system crashes in extreme cases. This unpredictability can disrupt the development and deployment of AI models, leading to missed deadlines, increased costs, and potential business losses.

AI systems process large volumes of data, which might include personal information, financial data, proprietary information, and other highly sensitive data. Protecting this data from unauthorized access, theft, or loss is a major concern. This protection is particularly challenging in the context of AI because the data often needs to be shared among various teams, systems, and possibly even locations, which can increase the risk of a data breach.

Introducing NetApp AIPOD™

Introducing NetApp® AIPOD™, an AI-focused converged infrastructure built on NetApp ONTAP® all-flash storage and powered by NVIDIA accelerated computing. Supported by the NVIDIA AI Enterprise software stack, NetApp BlueXP®, NetApp AI Control Plane, and NetApp DataOps Toolkit, AIPOD enables organizations to integrate with MLOps platforms and internal processes. NetApp AIPOD is a comprehensive AI solution designed to address the challenges that organizations face in adopting and implementing AI.

A result of the collaboration between NetApp and NVIDIA, NetApp AIPOD combines the strengths of both companies to offer a robust, scalable, and efficient platform for AI deployment. NetApp's hybrid cloud-enabled storage arrays are known for their high performance, scalability, and manageability. They're designed to handle the vast and diverse data that fuels AI, providing organizations with an effective strategy and solution for managing their data and executing AI wherever data exists. The NVIDIA AI platform delivers a full-stack solution that includes pretrained models, optimized software, and accelerated computing infrastructure solutions to enable organizations to reap a faster return on their AI investment.



NetApp AIPOD is a turnkey solution that addresses the challenges of AI adoption, offering a cost-effective, efficient, and scalable platform for AI deployment. It empowers organizations to fully harness the potential of AI, helping them to increase productivity, improve data governance, and gain a competitive advantage.

Data management simplicity

NetApp AIPOD significantly simplifies data management by offering an integrated, scalable solution that unifies AI workloads. Its unified hybrid data architecture facilitates the management of diverse data across different platforms and technologies, supporting file, block, and object protocols across cloud or on-premises environments. This architecture streamlines the management of data in various formats and across different platforms, reducing complexity. NetApp's data mobility capabilities enable enterprises to move data easily to meet AI practitioners where they are, exactly when they need it.

NetApp AIPOD is designed to handle vast, diverse data efficiently, and its flexible architecture adapts swiftly to changing data needs. Integration with NVIDIA's high-performance GPUs, purpose-built AI architectures, and NetApp AFF A-Series and C-Series arrays with NVIDIA GPU Direct Storage means powerful AI processing capabilities and rapid data access, maximizing the return on AI hardware investment.

With its DataOps Toolkit, NetApp AIPOD enables teams to automate AI training workflows and simplify provisioning, reducing the time and effort required to deploy AI models and accelerating insights. Built on NetApp's scalable storage solutions, AIPOD adapts quickly and efficiently to changing data needs, helping businesses grow their AI initiatives without the constraints of traditional siloed architectures.

And finally, the integration of AIPOd with NVIDIA's AI Enterprise software suite provides the necessary building blocks to develop and rapidly deploy end-to-end AI workflows, streamlining AI operations and reducing the complexity of integrating disparate systems and technologies.

High-performing data

NetApp AIPOd delivers high-performance data for AI workloads through its all-flash hybrid cloud-enabled storage arrays, designed to efficiently handle the data that fuels AI. The resulting rapid processing and analysis lead to efficient business operations. The flexible architecture of the NetApp AIPOd system scales efficiently, adapting to changing data needs that are crucial for growing AI initiatives.

NetApp AIPOd provides a unified hybrid data architecture, facilitating seamless integration of diverse platforms and technologies across cloud or on-premises environments. This solution breaks down barriers created by traditional, siloed architectures by offering a common platform that meets the unique requirements of different teams involved in AI initiatives.

Integration with NVIDIA's high-performance GPUs and reference architectures provides powerful AI processing capabilities for rapid training and deployment of AI models without performance limitations. NVIDIA's AI Enterprise software suite reduces the complexity of AI operations, resulting in rapid deployment of end-to-end AI workflows and adaptability to meet changing data needs. NetApp AIPOd eliminates storage-related bottlenecks through NetApp AFF A-Series and C-Series arrays with NVIDIA Magnum IO GPUDirect Storage. Additionally, the DataOps Toolkit enables the automation of AI training workflows and simplifies provisioning, so NetApp AIPOd accelerates AI model development and deployment, resulting in faster insights and improved business outcomes.

The most secure data

NetApp AIPOd equips enterprises to manage and process their most secure data safely and effectively through its unified hybrid data architecture, which includes robust data services for policy-driven governance and security. This architecture safeguards data, supports compliance efforts, and fosters transparency.

NetApp AIPOd offers an integrated data pipeline that intelligently manages data from edge to core to cloud for secure control and protection of data at all stages. It supports a unified hybrid data architecture with robust data services, enabling policy-driven governance and security, while automatically creating auditable copies of AI models and associated data for traceability and performance tracking. NetApp's advanced data management features, including Snapshot™ copies, FlexClone® files, and SnapLock® compliance software, provide a comprehensive audit trail for regulatory compliance, enhanced data governance, and protection against potential threats such as data loss, corruption, and ransomware.

NetApp and NVIDIA: Better together

NetApp and NVIDIA have a proven collaboration of 6 years. Together, they offer a field-tested, reliable AI infrastructure standard that is validated by thousands of deployments. NetApp is the only storage company that offers first-party services in the three major hyperscalers. NetApp's do-it-all data fabric for NVIDIA AI infrastructure delivers a hybrid cloud that enables insights up to 5 times faster. We know the challenges and have years of experience building tailor-made, validated AI solutions to solve them.



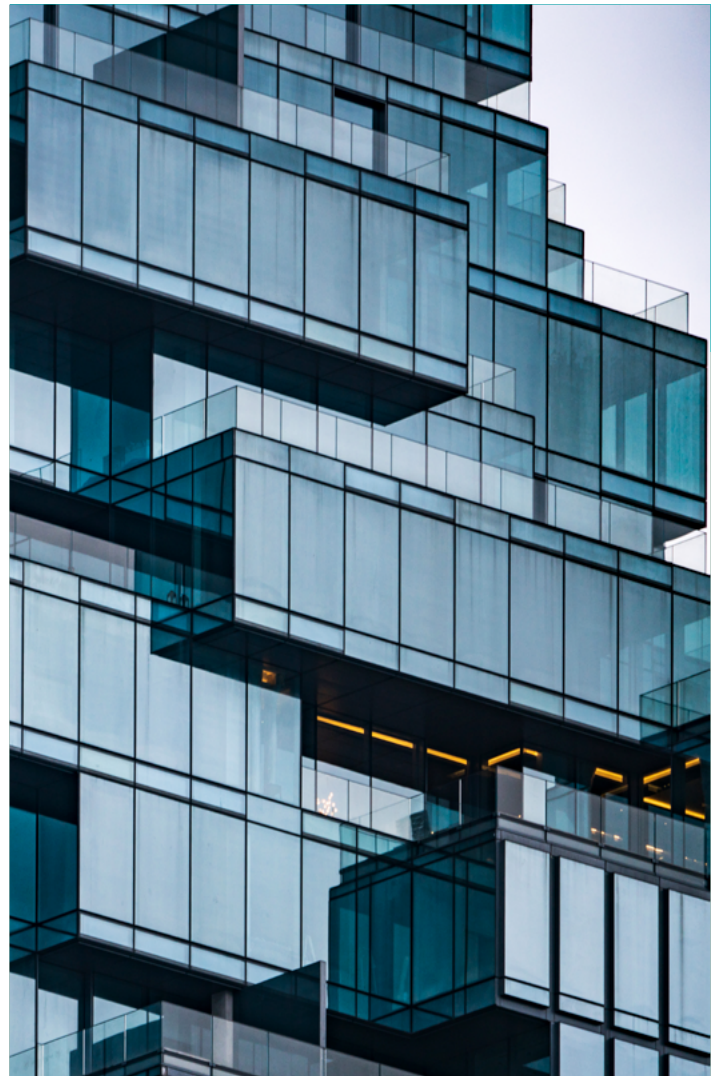
We offer a full spectrum of IT infrastructure consumption models, integrated and custom fit for all organizations. Hundreds of real-world deployments deliver real-world return on investment. ROI is optimized as AI moves into production environments, where performance and value count the most. Our results speak for themselves.

Conclusion

AI adoption presents great opportunities for enterprises of all sizes, but it also comes with significant challenges, including complex integrations, siloed infrastructures, unpredictable performance, scaling difficulties, and data security and management issues.

NetApp AIPOD, the result of a collaboration between NetApp and NVIDIA, is a comprehensive AI solution that addresses these challenges, offering a cost-effective, efficient, and scalable platform for AI deployment. It empowers organizations to fully harness the potential of AI, enabling them to increase productivity, improve data governance, and gain a competitive advantage.

With NetApp AIPOD, businesses can overcome the obstacles associated with AI adoption and accelerate their AI transformation journey, enabling them to revolutionize their operations, drive competitive advantage, and deliver superior solutions to customers.



About NetApp

NetApp is the intelligent data infrastructure company combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, then harnesses observability and AI, to enable the best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility and our data services create a data advantage through superior cyber-resilience, governance, and applications agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload or environment, transform your data infrastructure to realize your business possibilities with NetApp.

